Metabarcoding and Metagenomics 6: 227–238
DOI 10.3897/mbmg.6.85213

Research Article

The potential of metabarcoding plant components of Malaise trap samples to enhance knowledge of plant-insect interactions

Stephanie J. Swenson[1], **Lisa Eichler**[2], **Thomas Hörren**[3], **Andreas Kolter**[1], **Sebastian Köthe**[4], **Gerlind U. C. Lehmann**[4,5], **Gotthard Meinel**[2], **Roland Mühlethaler**[4], **Martin Sorg**[3], **Birgit Gemeinholzer**[1]

1 *University Kassel, Heinrich-Plett-Strasse 40, 34132 Kassel, Germany*
2 *Leibniz Institute of Ecological Urban and Regional Development (IOER), Weberplatz 1, 01217 Dresden, Germany*
3 *Entomological Society Krefeld (EVK), Marktstrasse 159, 47798 Krefeld, Germany*
4 *Nature and Biodiversity Conservation Union (NABU), Charitéstrasse 3, 10117 Berlin, Germany*
5 *Evolutionary Ecology, Humboldt University Berlin, Invalidenstrasse 110, 10115 Berlin, Germany*

Corresponding author: Stephanie J. Swenson (stephanie.swenson@uni-kassel.de)

Academic editor: Chloe Robinson | **Received** 11 April 2022 | **Accepted** 7 July 2022 | **Published** 22 July 2022

Abstract

The worldwide rapid declines in insect and plant abundance and diversity that have occurred in the past decades have gained public attention and demand for political actions to counteract these declines are growing. Rapid large-scale biomonitoring can aid in observing these changes and provide information for decisions for land management and species protection. Malaise traps have long been used for insect sampling and when insects are captured in these traps, they carry traces of plants they have visited on the body surface or as digested food material in the gut contents. Metabarcoding offers a promising method for identifying these plant traces, providing insight into the plants with which insects are directly interacting at a given time. To test the efficacy of DNA metabarcoding with these sample types, 79 samples from 21 sites across Germany were analysed with the ITS2 barcode. This study, to our knowledge, is the first examination of metabarcoding plant DNA traces from Malaise trap samples. Here, we report on the feasibility of sequencing these sample types, analysis of the resulting taxa, the usage of cultivated plants by insects near nature conservancy areas and the detection of rare and neophyte species. Due to the frequency of contamination and false positive reads, isolation and PCR negative controls should be used in every reaction. Metabarcoding has advantages in efficiency and resolution over microscopic identification of pollen and is the only possible identification method for the other plant traces from Malaise traps and could provide a broad utility for future studies of plant-insect interactions.

Key Words

biomonitoring, DNA metabarcoding, insect decline, landscape change, nature conservation, plant-insect interactions

Introduction

Landscape level change and chemical input in agriculture are major contributors to the rapid level of decline in diversity and abundance of insects observed in recent decades (Uhler et al. 2021). These declines are echoed in plant diversity over the past ~ 60 years (Eichenberg et al. 2020) and the events are most certainly linked to some degree. These rapid changes pose several risks to environmental and human health (Samways et al. 2020; van der Sluijs 2020) and necessitate improved, rapid methods of biomonitoring in order to address areas most affected by plant and insect declines. These improvements could aid in development of best practices to alleviate these declines in natural ecosystems, while protecting economic and agricultural concerns.

Due to these declines and the fast pace of landscape changes, development of novel methods of monitoring

Copyright *Stephanie J. Swenson et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

flora in order to understand which plant resources are directly used by insects on a temporal scale, could lead to better pest management practices in agriculture, as well as land management decisions for scale and spacing of conservation areas. Malaise traps have been used for collection of flying insects for more than 80 years (Malaise 1937) and with higher intensity after Townes (1972) published his trap model and provide a good assessment of the flying insects in an area at a given time (Skvarla et al. 2021). When insects enter the bulk Malaise trap preservative, they carry on their bodies evidence of the environments they have been living in and the organisms they have been feeding on. Recently, pesticide residues from the ethanol of these collection bottles have been analysed and provide information about the contamination of insects with pollutants (Brühl et al. 2021). The ability to identify not only the insects present in the Malaise trap, but also the plant traces they carry, could elucidate direct plant-insect interactions in a given time and space. This would enhance knowledge from traditional vegetation surveys, when they are available for an area, by providing information not only of what plants are available, but of what plants are being directly used by insects. Strategically placed Malaise traps could provide information on insect foraging and travel to urban areas or agricultural land from undeveloped or protected land, with the detection of garden (i.e. non-native ornamentals) or crop species in traps placed on protected land. Additionally, identification of these traces could aid in detection of threatened plant species and encroachment of neophyte or invasive species into new areas.

In spite of this potential, traditional microscopy identification of the pollen contained in Malaise traps is extremely time-consuming and requires extensive training and identification of plant fragments or regurgitated food material is nearly impossible. However, advancing techniques in genetic identification of complex mixed species environmental samples could provide a potential resource to identify the plant components found in Malaise trap samples. Metabarcoding, using a short gene region, or barcode, for the identification of many taxa contained in a complex sample, has displayed a great potential for plant and pollen identification over the last decade (Valentini et al. 2010; Jørgensen et al. 2012; Kraaijeveld et al. 2014; Hawkins et al. 2015; Keller et al. 2015; Fahner et al. 2016; Gous et al. 2019; Coughlan et al. 2021). The application of metabarcoding to the plant components found in Malaise traps offers an exciting prospect for environmental monitoring of insects and plant interactions. However, plant metabarcoding is not without its limitations. Several studies, as well as reviews and perspective articles, have indicated that universal standards, like selection of a strong barcode marker, primer choice and establishment of meaningful reference databases for plant metabarcoding are currently missing (Deiner et al. 2017; Dormontt et al. 2018; Ruppert et al. 2019; Banchi et al. 2020; Kolter and Gemeinholzer 2020, 2021). In addition, careful quality assurance in laboratory analysis, utilisation of sterile techniques, detection limits of very small sample sizes and PCR biases must be considered (Bell et al. 2017; Krehenwinkel et al. 2017), positive and negative controls must be added at each stage of laboratory work to ensure sample integrity and bioinformatic filtering must be performed carefully to detect technical artifacts that occur independent of true biological variants (Deiner et al. 2017; Stapleton et al. 2022). A further complication for all metabarcoding studies is the interpretation of read abundances to actual taxonomic abundances and each step of the collection and laboratory processes can affect the read quantity returned from sequencing. Several studies have detected unequal ratios between read and sample proportions or emphasised caution in interpretation of read abundance data and advocated for the inclusion of mock communities into study design (Albrecht and Leese 2015; Braukmann et al. 2019; Swenson and Gemeinholzer 2021). However, increasing studies have found metabarcoding to be semi-quantitative especially with the predominant taxa in a sample and there does appear to be some correlation between sample proportion and read number (Deagle et al. 2019; Polling et al. 2021).

In addition to the challenges inherent in plant barcoding and metabarcoding, the sample type presented by Malaise traps differs from those of previous studies in that the sample contains two signal types: 1) pollen and plant material as well as eDNA carried in on the insect body and 2) partially digested plant and pollen material excreted from the digestive tract or released due to breakage after capture. It is uncertain how these signals may interfere with each other and bias results.

This study aims to address whether Illumina MiSeq metabarcoding of plant fragments and eDNA found in Malaise trap preservative ethanol using the ITS2 barcode can retrieve a realistic assemblage of the vegetation available and utilised by the insects found in the traps. In addition, we examine whether rare or neophyte species can be detected in the samples and whether crop and non-native garden ornamentals are found in traps internal to protected areas indicating travel out of protected area for foraging. To evaluate our results, we developed the following hypotheses: 1) taxa retrieved will represent a realistic assemblage of German native plant taxa and complement the vegetation surveys taken in the sampling area; 2) plant species retrieved will primarily be those with pollen available in the respective regions and sampling duration; 3) insects will travel into and out of protected areas to forage and evidence of non-native garden plant and crop plant species will be found in the internal-most trap in the nature conservancy areas and 4) threatened and endangered plant species will be detected in the samples, but likely in very low read quantities.

Materials and methods

Collection methods/sites

As part of the project DINA (Diversity of Insects in Nature Protected Areas) (Lehmann et al. 2021) the

Entomological Society Krefeld (EVK), in cooperation with local farmers and nature conservation volunteers – mainly from the Nature and Biodiversity Conservation Union (NABU), established and maintained Malaise traps at 21 sites throughout Germany (Fig. 1) that encompassed the entire geographical area and a variety of habitats and each site was bordered by agricultural land (Lehmann et al. 2021).

At each of the 21 sites, five Malaise traps were placed at a gradient with the first located 25 m into arable land or as close as possible when the landscape would not permit. Subsequent traps were located 25 m distance from the other, with the second trap located directly on the intersection of arable and protected land and the fifth located 75 m towards the centre of the nature protected area (Lehmann et al. 2021). The nature protected areas meet the definition of International Union of Conservation of Nature (IUCN) Category IV habitat and species management, that aims to protect particular species or habitats, but vary in size and are not strictly prohibited from human use and often require management intervention. Vegetation surveys, including mosses and lichens, were conducted near the traps in 3.50 m^2 quadrats (Hallmann et al. 2017; Ssymank et al. 2018); however, mosses and lichens were not used in evaluations of metabarcoding. In addition to the quadrats, species occurring within a radius of 50 m around Malaise traps and the crops where the first trap was placed were recorded. Crops planted in the location of the first traps were two corn, wheat, and ryegrass fields, three rye and mixed cereal fields, as well as fallow land and one each of vineyard and peas.

Insects were collected using the standardised sampling design of German long-term studies of insect biomass (Hallmann et al. 2017) in 96% ethanol with a sampling interval of 14 days. From the continuous sampling from May through October, we selected a two-week period from the middle to end of May 2020, with a maximum difference in collection dates of four days across all sites, for this analysis. Following collection, EVK removed the original ethanol for plant metabarcoding. As the sample processing could not take place in a sterile environment, air pollen samples in the working space were collected to account for contamination during processing via the gravimetric method using a Petri dish (10 cm in diameter) coated with a thin layer of Vaseline. These samples were refrigerated until integration into the plant metabarcoding laboratory process. Insects were stored in new ethanol for voucher specimens, insect metabarcoding and morphological identification, although procedures are ongoing.

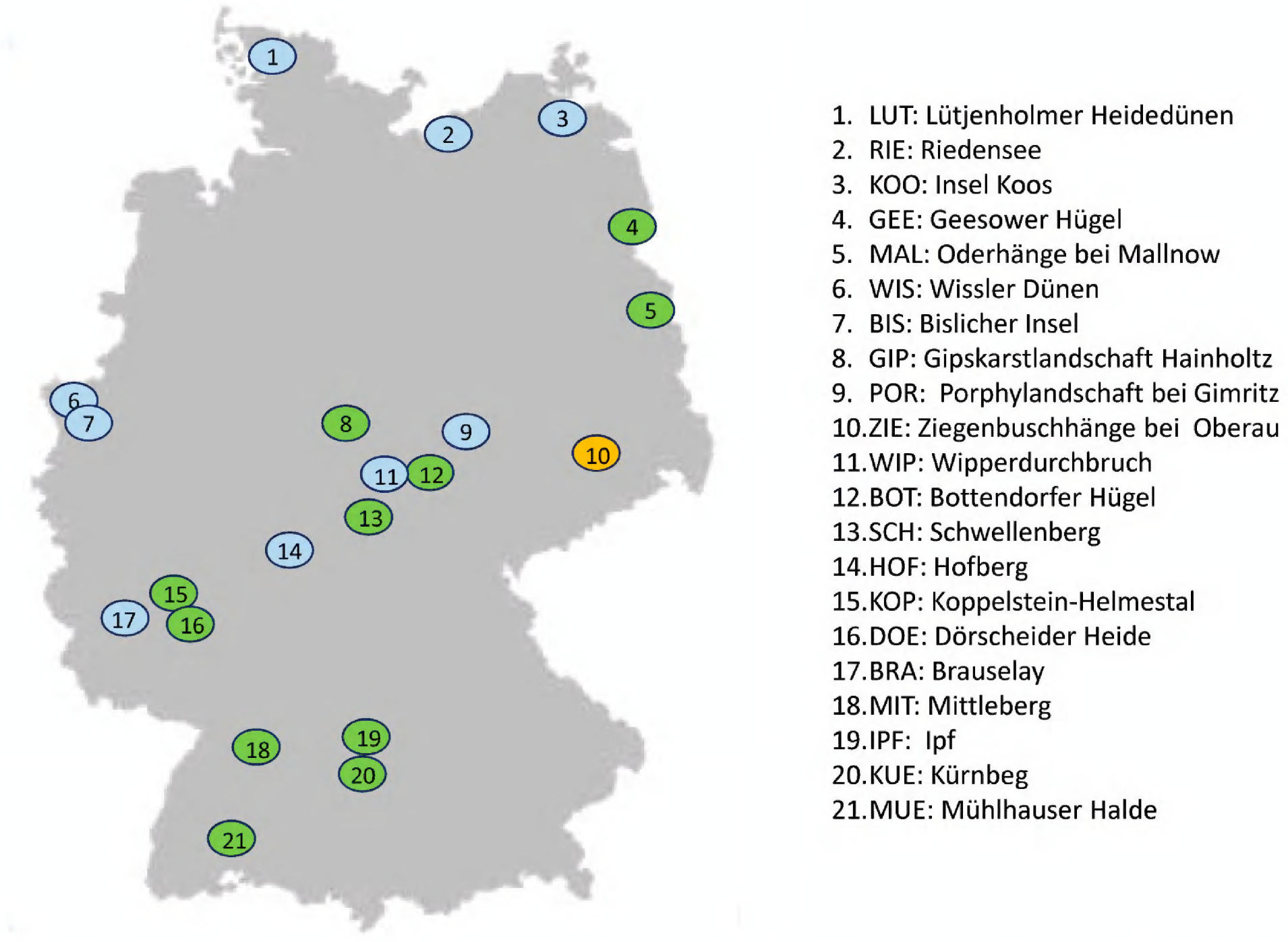

Figure 1. Map of the Malaise trap sampling sites throughout Germany, with site codes and names. Green indicates sites with complete data for all five traps, blue sites are missing data from at least one trap and orange indicates no data for the site.

Plant metabarcoding lab protocol

The original sample ethanol (200 ml ± 50 ml) was vacuum filtered using a 250 ml Nalgene single use analytical filter funnel with a cellulose nitrate (CN) filter (diameter 47 mm and 0.2 µl pore size) in a biosafety cabinet with DNA free equipment. Following filtration, the CN filter was cut into two equal parts and each part placed in a 2 ml Safe-Seal microcentrifuge tube (Sarstedt AG & Co. KG), with one half used for DNA extraction and the other saved as a voucher and/or backup for protocol optimisation. These samples were stored at -20 °C until further processing.

DNA extraction was performed with NucleoMag 96 Plant Kit (Macherey Nagel, Oesingen, Switzerland) with the following changes to the standard protocol: 1) 1 gm of 1.4 mm ceramic beads, 500 µl lysis buffer MC1, 5 µl Proteinase K (Macherey Nagel, Oesingen, Switzerland), 5 µl RNaseA were added to the 2 ml microcentrifuge tube containing the half filter paper with sediment and tissue was disrupted for 2.5 minutes with a Retsch MM400 bead mill at 30 Hz.; 2) following homogenisation, samples were incubated at 65 °C for one hour with constant shaking in addition to manual inversion mixing of the tubes every ten minutes to ensure uniformity of sample lysis; 3) following incubation, samples were centrifuged for ten minutes and the resulting 250–300 µl of lysate were transferred to clean 2 ml tubes; 4) 300 µl of binding buffer MC2 and 15 µl of magnetic beads were added; 5) remaining reagents were used at 25% of the standard protocol, with the exception of the elution buffer MC6 of which 35 µl were added and incubated at 50 °C for five minutes to evaporate any residual ethanol then 6) 25 µl were removed for PCR and sequencing and 2 µl for DNA quantification with Qubit 4 fluorometer (Thermo Fisher Scientific Inc.).

The ITS2 barcode was chosen for its high rate of success for species level identification, as well as having amongst the most abundant reference sequences available in public DNA sequencing repositories (Kolter and Gemeinholzer 2020). We used the ITS2 primers, Forward: ITS-3p62plF1, ACBTRGTGTGAATTGCAGRATC and Reverse: ITS-4unR1, TCCTCCGCTTATTKATATGC, shown to be optimal for amplification of plant DNA, based on *in silico* and mock community experimentation (Kolter and Gemeinholzer 2021). PCR was performed with three replicates per sample. Negative controls were added with three DNA extraction blanks and three PCR blanks and three replicates of a four species (*Ambrosia artemisiifolia*, *Fagus sylvatica*, *Lilium longiflorum* and *Plantago lanceolata*) mock community positive control to validate the efficacy of laboratory protocols. An adaptation of the Canadian Centre for Barcoding PlatinumTaq Protocol (Ivanova and Grainger 2007) was used for PCR with the addition of 0.25 µl of BSA (10 ng/µl) and 1.25 µl of 50% DMSO in a total volume of 12.5 µl per reaction. PCR cycling conditions were 95 °C for three minutes, followed by 35 cycles of 95 °C for 30 seconds, 50 °C for 30 seconds, 72 °C for 45 seconds, with a final extension of 72 °C for 10 minutes. Following PCR cycling, 5 µl of each of the three replicates were combined to produce a total volume of 15 µl and purified with Thermo Scientific Exonuclease 1. The pooled replicates of non-indexed PCR products were sent to LGC Genomics GmbH (Berlin) for sequencing on a MiSeq (2 × 300 bp) after an additional 12 PCR cycles. This additional PCR cycling consisted of three cycles at low annealing temperature (15 seconds 96 °C, 30 seconds 50 °C, 90 seconds 70 °C), followed by 9 cycles with increased annealing temperature (15 seconds 96 °C, 30 seconds 58 °C, 90 seconds 70 °C) with MyTaq Red Mix polymerase (Bioline BIO-25044).

Plant metabarcoding data pipeline

Sequencing data were processed with USEARCH (Edgar 2010) and DADA2 (Callahan et al. 2016) using R (v. 4.1.0) (R Core Team 2021). Prior to merging read pairs, sequencing primers were first trimmed, then quality filtered with a maximum expected error of 1.0 in USEARCH v.11. Following quality filtering, DADA2 was used to learn error rates and then denoised by the error profile with the pseudo-pooling function. Forward and reverse reads were merged with DADA2. Chimeras from denoised and merged read pairs were removed with Uchime3. The SINTAX algorithm (Edgar 2016) was used (cutoff 0.8) for ASV identification with a custom database (Suppl. material 2, 3).

The database was created in June 2021 from sequences downloaded using the GenBank webinterface at https://www.ncbi.nlm.nih.gov/genbank/ in GenBank (full) format using the search string (internal transcribed spacer1[Title] OR internal transcribed spacer 2[Title] OR ITS1[Title] OR ITS2[Title]) NOT patent NOT pseudogene NOT mRNA NOT unverified AND 100:2500[Sequence Length] AND Tracheophyta[Organism]. The term unverified is used by GenBank staff to flag erroneous sequences and should be excluded in every GenBank query. Due to the size of the downloaded data file, it was loaded into R by using the function fread from the package data.table. Taxonomic classifications and sequence information were extracted per GenBank accession number in the database. Species names were cleaned by removing any subspecies or variety information. Subsequently, unclassified environmental sequences and sequences containing any of the following characters [!§$%&/()=?`*'+#;`]., except for hyphen and underscore (the internally used string separator), were removed. Other irregularities which were removed include the presence of the keyword Eukaryota at the family descriptor and names starting with x_ or ending with a number. All plant family names were compared to the taxonomic backbone of Global Biodiversity Information Facility (GBIF) and discarded if no match with the status keyword ACCEPTED were found. The subsequent step removed all sequences with more than 1% ambiguous nucleotides. The ITSx algorithm was set to determine the stop and start positions of ITS1 and ITS2 sequences which matched a predefined

Tracheophyte HMMER profile, also supplied by ITSx (Bengtsson-Palme et al. 2013). All regions which were detected were extracted from the sequences and filtered by length. The length cut-off, for both regions individually, was set to a minimum of 100 and to a maximum of 400. The number of sequences per species was restricted to 10, with the longest sequences being favourably selected. To further harmonise the taxonomy, classifications higher than family level were completely replaced by the GBIF backbone taxonomy (GBIF Secretariat 2021).

ASVs with ambiguous species level identifications were given only genus level identifications. Taxa that do not occur in Germany and were likely the result of laboratory contamination were removed from analyses. Fungal contaminants were confirmed with BLAST search and removed. Values of Amplicon Sequence Variants (ASVs) found in extraction and PCR blanks were used to establish a relative abundance per negative threshold and ASVs not occurring above the highest relative abundance were removed. ASVs with less than five reads were removed.

Results

Sample evaluation

DNA yield of all samples (105) tended to be low with a range of 0.152 to 13.700 ng/µl (mean = 1.41, s.d. = 1.69). The majority of samples with a DNA yield below 0.5 ng/µl failed to amplify or produced low read abundances and were removed from analysis. Of the 21 sites, 10 sites produced full sequencing data for all five Malaise trap samples, 10 sites had 1–4 traps fail to amplify and one site had all traps failing to amplify. Attempts to amplify failing samples with different conditions have not yet been successful. Overall, 78 of the 105 samples provided sequencing data. The four species mock community positive controls retrieved all species and these species were absent in the Malaise trap samples. Extraction blanks and PCR blanks indicate a very low level of cross contamination. Plant reads were only rarely present (1, 3 and 90 reads in the three replications) in PCR blanks. Plant reads were more common in extraction blanks; however, they were limited to the most common taxa in a low read number (≤ 30) or were taxa only found in the extraction blanks (*Fraxinus excelsior* and *Cucumis sativus*).

Plant Identification

We identified 60 plant families in our reads, with 223 genus level identifications and 243 species level identifications (Suppl. material 1: Table S1), the majority of which are common throughout Germany and expected to be in the area of our sites. The range of different families in a site was 17 at Bislicher Insel (BIS) to 34 at Hofberg and Mühlhauser Halde (MUE, HOF) (Fig. 1). The dominant families by percentage of total reads were Brassicaceae (34.5%, with *Brassica* spp. at 29.3%), Fabaceae (9%), Rosaceae (9%), Poaceae (8.8%), Ranunculaceae (7.8%) and Pinaceae (6.5%) and one species in Adoxaceae, *Sambucus nigra* (4%).

The vegetation surveys recovered 48 flowering plant families, three of which were not recovered from metabarcoding (Orchidaceae, Hypericaceae and Linaceae). When extended to genera present, 77 were recovered only from vegetation surveys, 104 were recovered only from metabarcoding and 119 were recovered from both metabarcoding and vegetation surveys (Suppl. material 2: Table S2).

Species level identification of *Brassica* spp. ASVs was not possible due to their hybridogenous and polyploid origin. Based on the potential presence of *B. napus*, *B. nigra*, *B. oleracea*, and *B. rapa* in the sampling area and the hybridogenous DNA of *Brassica* cultivars, all of these taxa are likely included in the *Brassica* spp. ASV umbrella. Due to the prevalence in agriculture throughout Germany and May flowering time, we expect *B. napus* to be the most abundant species in the areas surrounding our experimental sites.

The highest generic level diversity was displayed in Brassicaceae (29), Poaceae (27), Asteraceae (21), Fabaceae (17), Rosaceae (12) and Caryophyllaceae (10). Of the taxa that could be assigned to at least generic level, only six were not likely to not have pollen available in the sampling time, the late summer or early autumn flowering *Hedera helix* and *Helianthus annus* and the late winter or early spring flowering *Alnus* spp., *Carpinus betulus*, *Corylus* spp. and *Taxus* spp.

Agricultural and garden plants in Malaise traps most internal to nature protected areas

We retrieved data for 14 Malaise traps placed most internal to the nature conservation area of the 21 sites. In these traps, we detected 21 agricultural or garden plants (Table 1). *Brassica* spp. detected in 12 sites and often in extremely high proportions of reads (> 90% Riedensee and Bottendorfer Hügel, > 70% Insel Koos, > 30% Geesower Hügel). Agricultural species of Poaceae were detected in a large proportion of sites (*Secale cereale* nine sites, *Triticum* spp. seven sites), but often at low read counts that could represent false positives. Eighteen species of garden plants, those not native to Germany, but commonly planted as ornamentals, were identified in internal traps. These were most often represented in low read counts which could represent false positives or are indicative of the low availability of garden plants as a foraging source.

Plant diversity detected in the traps at all sites combined

Red List and neophyte taxa detection

We detected 22 species listed as threatened (Metzing et al. 2018) in our samples (Table 2), with three species classified as highly threatened (Red List 2), six as threatened (Red List 3) and 13 as near threated (Red List V)

Table 1. Garden and Agricultural plants found in Malaise traps located most internal to nature conservancy areas with numbers indicate percentage of read abundance of the species in the sample.

Taxonomy			Sites												
Family	**Genus**	**Species**	**BOT**	**DOE**	**GEE**	**GIP**	**HOF**	**IPF**	**KOO**	**KOP**	**KUE**	**MAL**	**MIT**	**MUE**	**RIE**	**SCH**
Garden plants																
Adoxaceae	*Viburnum*	*opulus*	0	0	0	0	0	0.04	0	0	0.03	0	0	0	0	0
Amaryllidaceae	*Allium*	*ursinum*	0	0	0	0.43	0	0	0	0	0	0	0	0	0	0
Apiaceae	*Chaerophyllum*	*roseum*	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0
Apiaceae	*Heracleum*	*dissectum*	0	0	0	0	0	0	0	0	0	0	0.05	<0.01	0	0
Asteraceae	*Achillea*	*biebersteinii*	0	0.16	0	0	0	0	0	0	0	0	0	0	0	0
Asteraceae	*Helianthus*	*annuus*	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0
Asteraceae	*Pilosella*	*castellana*	0.53	0	0	0	0	0	0	0	5.39	0	0	0.30	0	0
Asteraceae	*Symphyotrichum*	*cordifolium*	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Brassicaceae	*Aubrieta*	*olympica*	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0
Brassicaceae	*Aubrieta*	sp.	0.01	0	0	0	0	0	0	0	0	0	0	0.70	0	0
Brassicaceae	*Aurinia*	*saxatilis*	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01
Caryophyllaceae	*Cerastium*	*alpinum*	0	0	0	0	<0.01	0	0	0	0.12	<0.01	0	0	0	<0.01
Cyperaceae	*Cyperus*	*diandrus*	0	0.06	0	0	0	0	0	0	0	0	0	0.02	0	0
Ericaceae	*Erica*	*arborea*	0	0	0	0	0	0	0	0	0	0	0.14	0	0	0
Fabaceae	*Wisteria*	sp.	0	0	0	0	0	0.10	0	0	0	0	0	0	0	0
Oleaceae	*Syringa*	*vulgaris*	0	0	0.02	0	0	0	0	0	0	0	0	<0.01	0	0
Poaceae	*Poa*	*trivialis*	0	0.03	0	<0.01	0.02	0.14	0	0.13	0.17	0.03	4.87	0.08	<0.01	0
Solanaceae	*Solanum*	*lycopersicum*	0	0	0	0	0	0	0	0	<0.01	0	0	<0.01	0	0
Agricultural plants																
Brassicaceae	*Brassica*	sp.	96.26	18.45	36.36	9.01	2.16	0.53	71.03	0	27.22	2.31	0	0.05	99.33	4.79
Poaceae	*Secale*	*cereale*	0.04	0.10	0.03	0	0.03	0.24	0	<0.01	0	2.41	0	0	<0.01	0.02
Poaceae	*Triticum*	sp.	<0.01	0.20	0	0	0.05	0.02	0	0	<0.01	0	0	0.02	0	0.02

according to the German Red List of plants (Metzing et al. 2018). However, these were often represented by low read counts (< 10) and could represent false positives. Thirteen neophyte species (Jäger et al. 2013) were detected in the samples and, as with the Red List species, the low read counts at certain sites could indicate false positives.

Discussion

Evaluation of methods

The low number of reads of taxa in extraction and PCR negative controls that also appear in the Malaise trap samples indicate a very low level of contamination between samples and soundness of our laboratory protocols. Additionally, the presence of all four species, contained in the positive control samples while not appearing in the Malaise trap samples, adds evidence of sound methods of extraction, primer choice and PCR protocols. Nevertheless 27 of our 105 samples failed to amplify, were primarily non-target fungal species or produced a low quantity read abundance. There are several possible reasons for these failures, including low quantity DNA extraction yield, impurities in the trap ethanol, variations in ethanol concentration, storage time until filtration and mechanical interference from non-plant debris found in the sample (Hallmaier-Wacker et al. 2018; Baksay et al. 2020; Kolter and Gemeinholzer 2021). This displays a need for optimisation of DNA extraction and/or PCR, evaluation of PCR inhibiting content, whether plant content in the sample is too low for downstream processing and evaluation of high ratios of non-plant material to plant material in the sample (i.e. Lepidoptera wing scales) that might disrupt performance. Our results indicate that there might not be a "one size fits all" method for samples of this type that vary in collection time or duration and environment and landscape type. A trade-off between obtaining complete data across all samples and high-throughput of a large number of samples may be necessary when designing a study including metabarcoding of Malaise trap plant components.

Taxonomic retrieval

Our samples recovered 223 genus level identifications and 243 species level identifications from 60 families. Diversity of species in our study was reflective of the species diversity within German taxa. Families with high species diversity like Asteraceae, Brassicaceae, Caryophyllaceae, Fabaceae, Poaceae and Rosaceae retrieved the largest number of species, while those with lower diversity retrieved (Suppl. material 1: Table S1) very low species diversity. When we compare family level results from metabarcoding to the families found in the vegetation surveys conducted in the sampling areas, metabarcoding recovered 15 families not found in the vegetation surveys and likely reflect the wide foraging range of insects captured. Three families of flowering plants were found in the vegetation surveys, but not recovered from metabarcoding. These families are represented by four species, three of which are very unlikely to bloom in May, *Hypericum maculatum* and *Hypericum perforatum*, with a blooming time of July to August and *Linum catharticum* with a blooming time of June to July and an orchid, *Gymnadenia conopsea*, with a specialised pollination system in which two pollen packages are dispersed on the head of one visiting

Table 2. Red list and neophyte plant species detected from ITS2 metabarcoding reads of Malaise trap plant components. *Indicates sites where the quantity of reads was low (≤10).

Status	Species	Number of sites	Site locations
Red list V	*Aira coaryphyllea*	1	DOE
Red list 3	*Alyssum Alyssoides*	1	SCH
Red list V	*Camelina microcarpa*	1	GEE
Red list V	*Camelina sativa*	1	GEE
Red list 2	*Chenopodiastrum murale*	1	POR*
Red list V	*Cynoglossum officinale*	3	GEE, POR, WIP
Red list V	*Eleocharis unglumis*	1	KOO*
Red list V	*Genista sagittalis*	2	DOE, WIP
Red list V	*Helictochloa sp.*	2	BOT, POR
Red list V	*Hippocrepis comosa*	5	HOF, IPF, KUE, MUE, WIP
Red list V	*Hottania palustris*	1	LUE
Red list V	*Lotus tenuis*	4	HOF, IPF*, KOP, MUE*
Red list 3	*Melamprum arvense*	1	GEE
Red list 3	*Myrica gale*	1	LUE
Red list 3	*Onobrychis viciifolia*	6	HOF, MAL*, MIT*, MUE, SCH, WIP*
Red list 2	*Papaver hybrium*	1	MAL
Red list V	*Primula veris*	3	GEE, HOF*, IPF*
Red list V	*Ranunculus polyanthemos*		GIP, KOP*, LUE*
Red list 2	*Saponaria ocymoides*	1	DOE*
Red list V	*Scleranthis perennis*	1	DOE*
Red list 3	*Silene conica*	1	MAL
Red list 3	*Silene otites*	1	SCH
Neophyte	*Anacyclus clavatus*	1	BOT
Neophyte	*Campanula portenschlagiana*	1	IPF
Neophyte	*Caragana* sp.	1	SCH*
Neophyte	*Hesperis* sp.	2	MUE, SCH
Neophyte	*Holcus annuus*	17	BIS*, BOT*, DOE, GIP, HOF*, IPF, KOO*, KOP, KUE*, LUE*, MAL, MIT, MUE, POR*, RIE, WIP*, WIS
Neophyte	*Lolium persicum*	1	WIS
Neophyte	*Medicago sativa*	6	DOE*, KUE*, MAL*, MIT*, SCH, WIS
Neophyte	*Monarda didyma*	1	MAL
Neophyte	*Pimpinella peregrina*	1	BRA
Neophyte	*Poa infirma*	3	LUE*, MAL*, RIE
Neophyte	*Robinia pseudoacacia*	9	BIS, BOT, BRA, DOE, IPF, KOP, MAL, MIT, POR,
Neophyte	*Trifolium incarnatum*	2	MAL, SCH
Neophyte	*Vicia pannonica*	1	SCH

insect making the chances of detection in Malaise traps extremely low (Suppl. material 2: Table S2).

The differences between the vegetation surveys and metabarcoding results are more pronounced when comparing at the generic level, where the number of genera recovered from both were 119, metabarcoding only 104 and vegetation surveys only 77. Ten tree genera were found only in metabarcoding and are likely to occur outside of the 50 m perimeter of the vegetation survey area.

The low diversity or absence of some taxa known to be in the area of the study sites can be attributed to several possible reasons. The first contributing factor is that Malaise traps are known to capture the largest proportions of the pollinator rich orders, Hymenoptera and Diptera (Prado et al. 2017; Karlsson et al. 2020; Brown 2021) and, for this reason, we expect the largest proportion of plants present in the sample to have pollen available during the sampling duration. This is reflected in our data where only four taxa (*Alnus* spp., *Carpinus betulus*, *Corylus* spp. and *Taxus* spp.) were present that are known to flower prior to May and one species, *Hedera helix*, that produces flowers from September to October. *Helianthus annuus* was also present at one site in a very low abundance (0.04%) and, while it is possible for this species to flower over our sampling time, it is typically a late summer flowering plant. Its presence at this site is more likely to be the result of sap feeding pests of sunflowers or a false positive. The majority of taxa we retrieved are known to or have the possibility of blossoming over the collection period, adding strength to our initial expectation of the assemblage of plant taxa we would recover from plant metabarcoding of these samples.

There are several other possible reasons for the low diversity or absence of certain taxa in our study sites beyond pollen availability. The first being the plants present in a sample represent insect preference and site-specific availability. While there are hundreds of species we did not recover from metabarcoding that may be plentiful in Germany as a whole, they might not be present in our study site or within the insect foraging range of our sites. In addition, some sites had already been managed with sheep grazing prior to this sampling duration and the Deutscher Wetterdienst (DWD) reported May 2020 to be very dry, both of these factors contributing to reduced plant resources in our study sites. After consideration of the time and space availability of particular plant species, recovery from metabarcoding is still dependent on the optimisation of DNA extraction, as well as its ability to overcome PCR bias. While we implemented good practice protocols of multiple PCR replicates, inclusion of positive controls and use of plant specific PCR primers that have been tested for optimal species recovery for this region, PCR bias will still have a role in the abundance and presence or absence of certain taxa within a sample (Aird et al. 2011; Elbrecht and Leese 2015).

Our utilisation of the ITS2 barcode could also partially explain the low level of retrieval of plants without available pollen. The plant DNA available from herbivory of vegetative plant structures will have been degraded by the digestion process and may require a barcode of shorter length to be amplified. Sample types composed of degraded DNA, such as herbal and food products, permafrost samples, faeces and ancient DNA, are more successfully resolved when a short DNA fragment, such as the P6 loop of *trnL* is implemented (Taberlet et al. 2007; Mallot et al. 2018; Gao et al. 2019). The *trnL*-P6 loop has been shown to recover more diversity on a family level when compared to ITS2 (Milla et al. 2021), but may be less informative by comparison at the genus and species level (Polling et al. 2021). Further testing is necessary, but the addition of the *trnL*-P6 loop could complement and enhance the results of plant metabarcoding from Malaise traps. This would be especially critical when the intent of the study is examination of non-pollinator plant-insect interactions.

Over half of our ASVs could not be identified to species level. Although the database used provided a high coverage for German taxa, estimated at 90%, based on growth and coverage of a previously used database for the German state of Bavaria created in 2014 (Sickel et al 2015), there are still missing species or species represented in very low levels of repetition. Kolter and Gemeinholzer (2021) demonstrated that addition of sequences of the same species to reference databases both increases correct species assignment and decreases erroneous identifications. Efforts to increase coverage of reference databases are a critical component to implementation of metabarcoding as a biomonitoring tool. However, even with the most optimal coverage and repetition of the reference database, there are taxa with complex evolutionary history that can likely never be identified to species level with metabarcoding. This is reflected in our results, where multiple genera such as *Brassica*, *Rubus* and *Taraxacum* could not be identified to species level, although we have knowledge of the species present in the study areas.

Occurrence of species of interest

Garden and agricultural plants

We detected several species of known garden plants and agricultural plants in the Malaise traps located most internally in the nature protected areas, indicating travel by a proportion of insects into urban and agricultural landscapes for foraging. Our results indicate a much larger foraging area than plants in the immediate vicinity of the Malaise traps, which is strengthened by the non-native ornamental garden plants present in our samples. Our study sites differ in their geographic vicinity to settlements and we do not have specific information on the plants grown in these areas, but it can be assumed they will offer a low resource for foraging compared to native plants in the area of the Malaise traps and the agricultural crops planted adjacent to them. The garden species represented in these samples are present in low read abundance and do not appear to be relevant to insect food availability in conservation areas at this time of year, but could be indicative of insect flight distances at times of lower pollen availability.

We also detected agricultural species, primarily *Brassica* spp., in the internal-most traps; however, the trap most internal in the nature protected area is 75 m from the edge of cultivated land, a distance that is within the recorded foraging range of most hymenopteran and dipteran pollinators (Zurbuchen et al. 2010; Inouye et al. 2015). In addition, the size and dimensions of many of the protected areas sites resulted in the most internal trap being very close to arable land from the opposite side. Locations of Malaise traps much further into large nature conservancy areas may be necessary to uncover the foraging behaviour of insects on protected and marginal land using metabarcoding.

The most prevalent agricultural group of plants in our samples were the *Brassica* spp. complex. The high level of *Brassica* spp. in the majority of samples confound data interpretation. The first problem being the complication in species level identifications from genetic methods due to the hybridogenous and polyploid origin within the genera. This limits the ability to use vegetation surveys to confirm real presence or absence with possible contamination from the field or laboratory processing or over-representation due to PCR bias. However, due to these species' (*B. napus*, *B. nigra*, *B. oleracea*, and *B. rapa*) prevalence in cultivation throughout Germany, as well as pollen of *B. napus* being an attractive and protein rich (Borutinskait et al. 2017) mass flowering food source, we believe that the abundance reflects true foraging events. This result indicates a need for further experimentation with samples when one species or group of species is expected to be dominant.

The other two agricultural species found in traps were Poaceae species, *Secale cereale* and *Triticum* spp. and, while they were present in the majority of traps, they were generally present with very low read numbers. Unlike the *Brassica* species, it seems more likely that the presence of these species was not the result of insect foraging. Poaceae is an anemophilic family and the Stiftung Deutscher Polleninformationsdienst (www.pollenstiftung.de) recorded this family to be in high pollen flight during the collection duration. The presence of these agricultural taxa in the samples, as well as other Poaceae, may be due to the ubiquitous occurrence of pollen in the environment where insects passively pick it up on their bodies. Some presence can also be explained by accidental, but unavoidable, contamination by practitioners in the field and laboratory. We detected several other taxa in our samples that were in peak pollen flight over our sampling duration, Pinaceae (6.5%), Poaceae (8.8%), *Rumex* spp. (2.3%). *Sambucus* spp. (4%), *Quercus* spp. (0.6%) and *Aesculus* spp. (0.7%). These results indicate a need for careful interpretation of results and incorporation of blank samples, as presence of a species recovered from metabarcoding might not represent a purposeful act by an insect and/or could be an accidental introduction from processing steps.

Red List species

While we did detect several threatened and near threatened Red List species (Metzing et al. 2018) in our samples (Table 2), our hypothesis that they would be rare in presence as well as read abundance was supported. This is an exciting prospect for biomonitoring the occurrence of rare plants; however, without careful interpretation of results and further corroboration from other data sources, great care must be taken in confirming whether presence of these species in sample represents a real presence in the environment. Additional experimentation should be performed to find detection limits for species of great concern and any incidence of threatened species in metabarcoding studies should be confirmed with vegetation surveys of the area before land management or protection decisions are made.

Neophyte detection

We detected 13 neophyte species in our samples, with generally higher read quantity and across more sites than the Red List species. The same caution in data interpretation and preliminary experimentation recommended for Red List species should be applied to detection of these taxa, especially as we do not know if the presence of these species in our samples occurred from plants present within or outside of the nature protected areas. Nevertheless, this result strengthens the potential of metabarcoding as a long-term biomonitoring tool for tracking and preventing the spread of deleterious plant species.

Overall plant taxa recovery

The results of metabarcoding for the most part were in line with the vegetation surveys when viewed from the family level and the three families not recovered from metabarcoding are likely due to flowering time or pollination strategy. Metabarcoding recovered 15 more families than found in the vegetation surveys, indicating plants that are not only found in the immediate sampling vicinity, but the greater insect foraging range. The dissimilarity of taxa is more pronounced when viewed from a generic level and, while some of this can be explained by the limitations of metabarcoding, it does not reach a level of failure of the methods. The majority of this dissimilarity in genera recovered can be explained by site specific availability, insect foraging range and flowering phenology. These results illustrate the complimentary nature of incorporation of metabarcoding into Malaise trap biomonitoring programmes, vegetation surveys can give the entire view of plants in the area, while metabarcoding could potentially enhance knowledge of what plants are being used at a given time and how these assemblages change over a growing season.

Conclusions

Our study illustrates the potential of Malaise trap plant metabarcoding as an additional tool for large-scale plant biomonitoring; however, cautious consideration of its limitations must be included in project design, data analysis and interpretation. Further experimentation must be undertaken to account for sample failure with mock communities and examination of several barcodes and primer combinations should be evaluated. When specific species are of great interest to the particular study, preliminary experimentation with mock communities must be conducted to determine detection limits of the species. Unintentional introduction of airborne species can greatly affect the relative read proportions retrieved and confound data interpretation, creation of novel blanks that could be added to field protocols could aid in accounting field and lab-introduced contamination, rather than the occurrence from insect interaction in the environment.

Funding

The Project DINA is funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF; grant number FKZ 01LC1901).

Competing interests

The authors have declared that no competing interests exist.

Data accessibility

Raw sequence data is available on the Sequence Read Archive (SRA) under the accession number PRJNA851235. The R script for the identification database, as well as the database, are in Suppl. material 3.

Acknowledgements

We are especially thankful to all the volunteers in the field who maintained the Malaise traps. Furthermore, we would like to thank all partners of the DINA consortium for the collaboration and scientific input and especially Wolfgang Wägele for project initiation. Thank you to Sabine Mutz, Kathrin Moses-Luehrsen and Annalena Kurzweil (all University Giessen) for laboratory and administrative support.

References

Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology 12(2): R18. https://doi.org/10.1186/gb-2011-12-2-r18

Baksay S, Pornon A, Burrus M, Mariette Andalo C, Excaravage N (2020) Experimental quantification of pollen with DNA metabarcoding using ITS1 and *trnL*. Scientific Reports 10(1): 4202. https://doi.org/10.1038/s41598-020-61198-6

Banchi E, Ametrano CG, Greco S, Stanković D, Muggia L, Pallavicini A (2020) Stanković, Muggia L, Pallavicini A (2020) PLANiTS: A curated sequence reference dataset for plant ITS DNA metabarcoding. Database 2020: baz155. https://doi.org/10.1093/database/baz155

Bell KL, Fowler J, Burgess KS, Dobbs EK, Gruenewald D, Lawley B, Morozumi C, Brosi BJ (2017) Applying pollen DNA metabarcoding to the study of plant-pollinator interactions. Applications in Plant Sciences 5(6): 1600124. https://doi.org/10.3732/apps.1600124

Bengtsson-Palme P, Veldre V, Ryberg M, Hartmann M, Branso S, Wang Z, Godhe A, Bertrand Y, De Wit P, Sanchez M, Ebersberger I, Sanli K, de Souza F, Kristiansson E, Abarenkov K, Eriksson KM, Nilsson RH (2013) ITSx: Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukary-

otes for use in environmental sequencing. Methods in Ecology and Evolution 4: 914–919. https://doi.org/10.1111/2041-210X.12073

Borutinskait V, Treigyt G, Matuzevičius D, Zaikova I, Ceksteryt V, Navakauskas D, Kurtinaitien B, Mayakovskian R (2017) Proteomic analysis of pollen and blossom honey from rape seed *Brassica napus* L. Journal of Apicultural Science 61(1): 73–92. https://doi.org/10.1515/jas-2017-0006

Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, de Waard JR, Sones JE, Zakharov EV, Hebert PDN (2019) Metabarcoding a diverse arthropod mock community. Molecular Ecology Resources 19(3): 711–727. https://doi.org/10.1111/1755-0998.13008

Brown BV (2021) Sampling methods for adult flies (Diptera). In: Santos JC, Fernandes GW (Eds) Measuring arthropod diversity. Springer, Cham, 187–204. https://doi.org/10.1007/978-3-030-53226-0_7

Brühl CA, Bakanov N, Köthe S, Eichler L, Sorg M, Hörren T, Mühlethaler R, Meinel G, Lehmann GU (2021) Direct pesticide exposure of insects in nature conservation areas in Germany. Scientific Reports 11(1): 1–10. https://doi.org/10.1038/s41598-021-03366-w

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Coughlan SA, Shafer ABA, Freeland JR (2021) Development of an environmental DNA metabarcoding assay for aquatic vascular plant communities. Environmental DNA 3(2): 372–387. https://doi.org/10.1002/edn3.120

Deagle BE, Thomas AC, McInnes JC, Clark LJ, Vesterinen EJ, Clare EL, Eveson JP (2019) Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? Molecular Ecology 28(2): 391–406. https://doi.org/10.1111/mec.14734

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursiére-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, DeVere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26: 5872–5895. https://doi.org/10.1111/mec.14350

Dormontt EE, van Dijk K, Bell KL, Biffin E, Breed MF, Byrne M, Caddy-Retalic S, Encinas-Viso F, Nevill PG, Shapcott A, Young JM, Waycott M, Lowe AJ (2018) Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections –An Australian perspective. Frontiers in Ecology and Evolution 13: 134. https://doi.org/10.3389/fevo.2018.00134

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19): 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar RC (2016) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. bioRxiv, 1–20. https://doi.org/10.1101/074161

Eichenberg D, Bowler DE, Bonn A, Bruelheide H, Grescho V, Harter D, Jandt U, May R, Winter M, Jansen F (2020) Widespread decline in central European plant diversity across six decades. Global Change Biology 27(5): 1097–1110. https://doi.org/10.1111/gcb.15447

Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass – Sequence relationships with an innovative metabarcoding protocol. PLoS ONE 10(7): e0130324. https://doi.org/10.1371/journal.pone.0130324

Fahner N, Shokralla S, Baird D, Hajibabaei M (2016) Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, Resolution, and annotation of four DNA markers. PLoS ONE 11(6): e015750. https://doi.org/10.1371/journal.pone.0157505

Gao Z, Liu Y, Wang X, Wei X, Han J (2019) DNA Mini-Barcoding: A derived barcoding method for herbal molecular identification. Frontiers in Plant Science 10: 987. https://doi.org/10.3389/fpls.2019.00987

Gous A, Swanevellder DZH, Eardley CD, Willows-Munro S (2019) Plant-pollinator interactions over time: Pollen metabarcoding from bees in a historic collection. Evolutionary Applications 12(12): 187–197. https://doi.org/10.1111/eva.12707

Hallmaier-Wacker LK, Lueert S, Roos C, Knauf S (2018) The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. Scientific Reports 8(1): 6292. https://doi.org/10.1038/s41598-018-24573-y

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hörren T, Goulson D, de Kroon H (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE 12(10): e0185809. https://doi.org/10.1371/journal.pone.0185809

Hawkins J, De Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, Baillie L, Adams-Groom B (2015) Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS ONE 10(8): 20. https://doi.org/10.1371/journal.pone.0134735

Inouye DW, Larson BMH, Ssymank A, Kevan PG (2015) Flies and flowers III: Ecology of foraging and pollination. Journal of Pollination Ecology 16(6): 115–133. https://doi.org/10.26786/1920-7603(2015)15

Ivanova NV, Grainger CM (2007) CCDB Protocols, COI amplification. https://ccdb.ca/site/wp-content/uploads/2016/09/CCDB_Amplification.pdf

Jäger EJ, Ritz CM, Müller F, Welk E, Wesche K (2013) Rothmaler – Exkursionsflora von Deutschland Teil: Gefäßpflanzen: Atlasband. Springer Sprektrum, Heidelberg, 822 pp. https://doi.org/10.1007/978-3-662-49708-1

Jørgensen T, Kjær KH, Haile J, Rasmussen M, Boessenkool S, Andersen K, Coissac E, Taberlet P, Brochman C, Orlando L, Gilbert TP, Willerslev E (2012) Islands in the ice: Detecting past vegetation on Greenlandic nuntaks using historical records and sedimentary ancient DNA Meta-barcoding. Molecular Ecology 21(8): 1980–1988. https://doi.org/10.1111/j.1365-294X.2011.05278.x

Karlsson D, Hartop E, Forshage M, Jaschhof M, Ronquist F (2020) The Swedish Malaise trap project: A 15-year retrospective on a country-wide insect inventory. Biodiversity Data Journal 8: e47255. https://doi.org/10.3897/BDJ.8.e47255

Keller A, Danner N, Ankenbrand M, von der Ohe K, von der Ohe W, Rost S, Härtel S, Steffan-Dewenter I (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biology 17(2): 558–566. https://doi.org/10.1111/plb.12251

Kolter A, Gemeinholzer B (2020) Plant DNA barcoding necessitates marker-specific efforts to establish more comprehensive reference databases. Genome 64(3): 265–298. https://doi.org/10.1139/gen-2019-0198

Kolter A, Gemeinholzer B (2021) Internal transcribed spacer evaluation for vascular plant metabarcoding. Metabarcoding and Metagenomics 5: 133–152. https://doi.org/10.3897/mbmg.5.68155

Kraaijeveld K, de Weger LA, Garcia MV, Buermans H, Frank J, Hiemstra PS, den Dunnen JT (2014) Efficient and sensitive identificatioin

and quantification of airborne pollen using next-generation DNA sequencing. Molecular Ecology Resources 15(1): 8–16. https://doi.org/10.1111/1755-0998.12288

Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG (2017) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports 7(1): 17668. https://doi.org/10.1038/s41598-017-17333-x

Lehmann GUC, Bakanov N, Behnisch M, Bourlat SJ, Brühl CA, Eichler L, Fickel T, Geiger MF, Gemeinholzer B, Hörren T, Köthe S, Lux A, Meinel G, Mühlethaler R, Poglitsch H, Schäffler L, Schlechtrimen U, Schneider FD, Schulte R, Sorg M, Sprenger M, Swenson SJ, Terlau W, Turck A, Zizka VMA (2021) Diversity of insects in nature protected areas (DINA): And interdisciplinary German research project. Biodiversity and Conservation 30(8–9): 2605–2614. https://doi.org/10.1007/s10531-021-02209-4

Malaise R (1937) A new insect trap. Entomologisk Tidskrift 58: 148–160.

Mallot EK, Garber PA, Malhi RS (2018) *trnL* outperforms *rbcL* as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (*Cebus capucinus*, Primates). PLoS ONE 13(6): e0199556. https://doi.org/10.1371/journal.pone.0199556

Metzing D, Hofbauer N, Ludwig G, Matzke-Hajek G (2018) Rote Liste gefährdeter Tiere, Pflanzen und Pilze Deutschlands - Band 7: Pflanzen. Bundesamt für Naturschutz 70/7.

Milla L, Sniderman K, Lines R, Mousavi-Derazmahalleh M, Encinas-Viso F (2021) Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. Ecology and Evolution 11(13): 8683–8698. https://doi.org/10.1002/ece3.7679

Polling M, Sin M, de Weger LA, Speksnijder AG, Koenders MJ, de Boer H, Gravendeel B (2021) DNA metabarcoding using nrITS2 provides highly quantitatve results for airborne pollen monitoring. The Science of the Total Environment 806: 150468. https://doi.org/10.1016/j.scitotenv.2021.150468

Prado SG, Ngo HT, Florez JA, Collazo JA (2017) Sampling bees in tropical forests and agroecosystems: A review. Journal of Insect Conservation 21(5–6): 753–770. https://doi.org/10.1007/s10841-017-0018-8

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. https://www.R-project.org/

Ruppert KM, Kline RJ, Rahman MS (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring and applications of global eDNA. Global Ecology and Conservation 17: e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Samways MJ, Barton PS, Birhofer K, Chichorro F, Deacon C, Fartmann T, Fukushima CS, Gaigher R, Habel JC, Hallmann CA, Hill MJ, Hochkirch A, Kaila L, Kwak ML, Maes K, Mammola S, Noriega JA, Orfinger AB, Pedraza F, Pryke JS, Roque FO, Settele J, Samake JP, Stork NE, Suhling F, Vorster C, Cardoso P (2020) Solutions for humanity on how to conserve insects. Biological Conservation 242: 108427. https://doi.org/10.1016/j.biocon.2020.108427

GBIF Secretariat (2021) GBIF Backbone Taxonomy. Checklist dataset. https://doi.org/10.15468/39omei

Sickel W, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, Steffan-Dewenter I, Keller A (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecology 15(20): 20. https://doi.org/10.1186/s12898-015-0051-y

Skvarla MJ, Larson JL, Fisher JR, Dowling SP (2021) A review of terrestrial and Canopy Malaise Traps. Annals of the Entomological Society of America 114(1): 27–47. https://doi.org/10.1093/aesa/saaa044

Ssymank A, Sorg M, Doczkal D, Rulik B, Merkel-Wallner G, Vischer-Leopold M (2018) Praktische Hinweise und Empfehlungen zur Anwendung von Malaise fallen für Insekten in der Biodiversitätserfassung und im Monitoring, 1–12.

Stapleton TE, Weinstein SB, Greehalgh R, Deering MD (2022) Successes and limitations of quantitative diet metabarcoding in a small herbivorous mammal. Molecular Ecology Resources 00: 1–14. https://doi.org/10.1111/1755-0998.13643

Swenson SJ, Gemeinholzer B (2021) Testing the effect of pollen exine rupture on metabarcoding with Illumina sequencing. PLoS ONE 16(2): e0245611. https://doi.org/10.1371/journal.pone.0245611

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. Nucleic Acids Research 35(3): e14. https://doi.org/10.1093/nar/gkl938

Townes H (1972) A light-weight Malaise trap. Entomological News 83: 239–247.

Uhler J, Redlich S, Zhang J, Hothorn T, Tobisch C, Ewald J, Thorn S, Seibold S, Mitesser O, Morinière J, Bozicevic V, Benjamin CS, Englemeier J, Fricke U, Ganuza C, Haensel M, Riebl R, Rojas-Botero S, Rummler T, Uphus L, Schmidt S, Steffan-Dewenter I, Müller J (2021) Relationship of insect biomass and richness with land use along a climate gradient. Nature Communications 12(1): 5946. https://doi.org/10.1038/s41467-021-26181-3

Valentini A, Miquel C, Taberlet P (2010) DNA barcoding for honey biodiversity. Diversity 2(4): 610–617. https://doi.org/10.3390/d2040610

van der Sluijs JP (2020) Insect decline, and emerging global environmental risk. Current Opinion in Environmental Sustainability 46: 39–42. https://doi.org/10.1016/j.cosust.2020.08.012

Zurbuchen A, Landert L, Klaiber J, Müller A, Hein S, Dorn S (2010) Maximum foraging ranges in solitary bees: Only few individuals have the capability to cover long foraging distances. Biological Conservation 143(3): 3. https://doi.org/10.1016/j.biocon.2009.12.003

Supplementary material 1

Table S1

Author: Stephanie J. Swenson, Lisa Eichler, Thomas Hörren, Andreas Kolter, Sebastian Köthe, Gerlind U. C. Lehmann, Gotthard Meinel, Roland Mühlethaler, Martin Sorg, Birgit Gemeinholzer

Data type: table

Explanation note: Generic and species level identification of Amplicon Sequence Variance (ASVs) resulting from ITS2 metabarcoding of Malaise trap plant contents.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.85213.suppl1

Supplementary material 2
Table S2

Author: Stephanie J. Swenson, Lisa Eichler, Thomas Hörren, Andreas Kolter, Sebastian Köthe, Gerlind U. C. Lehmann, Gotthard Meinel, Roland Mühlethaler, Martin Sorg, Birgit Gemeinholzer

Data type: table

Explanation note: Species retrieved from ITS2 metabarcoding and vegetation surveys in the area surrounding Malaise traps.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.85213.suppl2

Supplementary material 3
R script for ITS identification database

Author: Stephanie J. Swenson, Lisa Eichler, Thomas Hörren, Andreas Kolter, Sebastian Köthe, Gerlind U. C. Lehmann, Gotthard Meinel, Roland Mühlethaler, Martin Sorg, Birgit Gemeinholzer

Data type: R script

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.85213.suppl3